

**ANALISIS DAN IMPLEMENTASI PENDEKATAN KORELASI
UNTUK PERBAIKAN *MISSING VALUE* DENGAN METODE
K-NEAREST NEIGHBOR (K-NN)**

TESIS

**PROGRAM MAGISTER TEKNIK ELEKTRO
MINAT SISTEM KOMUNIKASI DAN INFORMATIKA**

Ditujukan untuk memenuhi persyaratan
memperoleh gelar Magister Teknik



NOVTA DANY'EL IRAWAN
NIM. 146060300111032

**UNIVERSITAS BRAWIJAYA
FAKULTAS TEKNIK
MALANG
2017**

TESIS

ANALISIS DAN IMPLEMENTASI PENDEKATAN KORELASI UNTUK
PERBAIKAN *MISSING VALUE* DENGAN METODE *K-NEAREST NEIGHBOR*

NOVTA DANY'EL IRAWAN

NIM. 146060300111032

telah dipertahankan di depan penguji
pada tanggal
dinyatakan telah memenuhi syarat
untuk memperoleh gelar Magister Teknik

Komisi Pembimbing,

Ketua,

Anggota,

Ir. Wijono, M.T., Ph.D
NIP. 196211111989031003

Dr-Ing. Onny Setyawati, S.T., M.T, M.Sc.
NIP. 197404172000032007

Malang,

Universitas Brawijaya
Fakultas Teknik, Jurusan Teknik Elektro
Ketua Program Studi Magister Teknik Elektro

Dr. Eng. Panca Mudjirahardjo ST., MT.
NIP. 197003292000121001

PERNYATAAN ORIGINALITAS TESIS

IDENTITAS TIM PENGUJI TESIS

JUDUL TESIS : ANALISIS DAN IMPLEMENTASI PENDEKATAN
KORELASI UNTUK PERBAIKAN *MISSING VALUE*
DENGAN METODE *K-NEAREST NEIGHBOR* (K-NN)

Nama Mahasiswa : Novta Dany'el Irawan
NIM : 146060300111032
Program Studi : Program Magister Teknik Elektro
Minat : Sistem Komunikasi dan Informatika

KOMISI PEMBIMBING

Ketua : Ir. Wijono M.T, Ph.D.
Anggota : Dr-Ing. Onny Setyawati, S.T., M.T., M.Sc.

TIM DOSEN PENGUJI

Penguji 1 : Dr. Ir. Erni Yudaningtyas, M.T.
Penguji 2 : Dr. M. Sarosa, Dipl.Ing., M.T.
Tanggal Ujian : 24 November 2017
SK Penguji : 1539 Tahun 2017

RIWAYAT HIDUP PENULIS



Novta Dany'el Irawan dilahirkan di Pasuruan pada tanggal 25 November 1989 sebagai anak Pertama dari dua bersaudara dari pasangan Bambang Irawan dan Murtiningsih. Pendidikan dimulai dari TK Dharmawanita Purwodadi, dilanjutkan Sekolah Dasar ditempuh di SDN - Purwodadi 1, kemudian melanjutkan pendidikannya di SMPN 1 - Purwosari dan, SMK plus Almaarif Jurusan Teknik Informatika - Singosari. Pendidikan Diploma (D3) di tempuh di Jurusan Teknik Elektro, Prodi Elektronika Universitas Negeri Malang, lulus pada tahun 2011. Dilanjutkan Pada jenjang Strata 1 (S1) Jurusan Teknik Elektro, Prodi Pend.Teknik Informatika di Universitas Negeri Malang, lulus tahun 2014. Pada tahun 2014 melanjutkan pendidikan Strata 2 (S2) di Universitas Brawijaya Malang, pada jurusan Magister Teknik Elektro, peminatan Sistem Komunikasi dan Informatika, dan menamatkannya pada tahun 2017.

KATA PENGANTAR

Puji syukur kami haturkan kehadiran Allah SWT, karena dengan karunia-Nya penulis dapat menyelesaikan Tesis yang berjudul “ Analisis Dan Implementasi Pendekatan Korelasi Untuk Perbaikan *Missing Value* Dengan Metode *K-Nearest Neighbor*”. *Missing value* sering terjadi dalam metode klasifikasi disebabkan karena informasi tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada, hal ini akan menyebabkan menurunnya keakuratan dan kualitas data pada saat data diolah. Hal inilah yang membuat topik ini sangat menarik, karena banyak permasalahan dalam pemrosesan klasifikasi yang terdapat *missing value* bisa diangkat sebagai *research problem* dalam sebuah penelitian.

Dalam menyelesaikan Tesis ini, penulis melaksanakan penelitian berdasarkan teori yang telah didapatkan dari perkuliahan, studi literatur, dan bimbingan dari dosen pembimbing. Penulis sadar bahwa masih banyak kekurangan dalam penulisan Tesis ini, untuk itu penulis mohon maaf dan mengharapkan kritik serta masukan guna pengembangan ilmu kedepan. Saran serta masukan dapat dikirim ke alamat *email*: novtadirawan@gmail.com. Semoga dikemudian hari, Tesis ini memberikan manfaat dalam pengembangan ilmu.

Malang,.....2017

Novta Dany’el Irawan

UCAPAN TERIMA KASIH

Penulis menyadari terlaksananya pembuatan Tesis dan terwujudnya karya ilmiah ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Karena itu pada kesempatan ini izinkanlah penulis menyampaikan ucapan terima kasih kepada:

1. Allah SWT yang telah memberikan kesehatan, kekuatan, kesabaran dan ilmu yang barokah sehingga penulis dapat menyelesaikan Tesis ini dengan baik.
2. Nabi Muhammad SAW yang menjadi suri tauladan bagi penulis.
3. Papa Bambang Irawan, Mama Murtiningsih, dan Saudariku Yulia Yustraeni Irawan tercinta atas segala pengertian, dukungan, bantuan, semangat dan doa yang diberikan pada penulis.
4. Seseorang paling spesial dalam hidup penulis Galih Dianing Fitri, S.Pd yang tidak pernah lelah memberi dukungan dan doa sehingga penulis dapat menyelesaikan Tesis ini.
5. Bapak Ir. Wijono M.T, PhD., dan Ibu Dr.-Ing Onny Setyawati S.T, M.T, M.Sc. selaku Pembimbing, terima kasih atas segala bimbingan, nasehat, kritikan dan dukungannya sehingga Tesis ini dapat diselesaikan dengan baik, semoga penulis bisa meneruskan ilmu yang didapatkan ini kepada orang lain.
6. Ibu Dr. Ir. Erni Yudaningtyas, M.T., dan Bapak Dr. M. Sarosa, Dipl.Ing., M.T. selaku penguji, terima kasih atas saran, kritik dan masukan yang membuat karya ilmiah ini menjadi lebih baik.
7. Seluruh dosen pengajar beserta staff di Program Magister dan Doktor Fakultas Teknik atas bimbingan, bantuan dan ilmu pengetahuan yang diberikan.
8. Keluarga besar Program Magister teknik Elektro Universitas Brawijaya, khususnya SKI C1 dan C2 2014 yang telah memberikan bantuan moril, semangat dan dorongan kepada penulis.

Dan semua pihak yang tidak bisa saya sebutkan satu persatu, yang secara langsung maupun tidak langsung turut membantu menyelesaikan Tesis ini, terima kasih atas segala dukungan dan doa yang diberikan kepada penulis, semoga Tuhan Yang Maha Esa melimpahkan rahmat dan hidayat sebagai imbalan atas kebaikan yang diberikan kepada penulis.

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	i
PERNYATAAN ORIGINALITAS	ii
IDENTITAS TIM PENGUJI TESIS	iii
RIWAYAT HIDUP PENULIS	iv
KATA PENGANTAR	v
UCAPAN TERIMA KASIH	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
RINGKASAN	xi
SUMMARY	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
BAB II KAJIAN PUSTAKA	
2.1 Hasil Penelitian Relevan	5
2.2 Tinjauan Teori	6
2.2.1 Tujuan Penjurusan Siswa	6
2.2.2 <i>Missing Value</i>	8
2.2.3 Sistem Pengambil Keputusan (SPK).....	8
2.2.4 Pendekatan Korelasi.....	9
2.2.5 Analisis Data pada Korelasi.....	10
2.2.6 Klasifikasi <i>K-Nearest Neighbor</i> (K-NN).....	11
BAB III KERANGKA KONSEP PENELITIAN	
3.1 Kerangka Teori Penelitian.....	13
3.2 Alur Langkah Penelitian.....	15
3.2.1 Pengumpulan Data	15

3.2.2 Pendekatan Korelasi.....	16
3.2.3 Perancangan.....	16
3.2.4 Klasifikasi.....	17
3.2.5 Pengujian	17
3.3. Hipotesis Penelitian	17
BAB IV METODOLOGI PENELITIAN	
4.1 Populasi dan Sampel.....	19
4.2 Persiapan Data dan Analisa Data	20
4.2.1 Pengambilan Data	20
4.2.2 Kusioner	21
4.2.3 Analisis Data.....	23
4.3 Klasifikasi <i>K-Nearest Neighbor</i>	26
4.4 Uji Akurasi Hasil Klasifikasi	27
BAB V HASIL DAN PEMBAHASAN	
5.1 Hasil Tingkat Korelasi Instrumen Cadangan	29
5.2 Klasifikasi Jenis Variabel Cadangan	31
5.3 Pola Kasus Lama	33
5.4 Klasifikasi dan Penggantian Variabel	34
5.4.1 Penggantian <i>Variable Missing Value</i> dengan Variabel Cadangan ..	34
5.4.2 Pengukuran Akurasi.....	38
5.4.3 Kasus Proses Klasifikasi K-NN.....	39
5.4 Analisis Hasil Penelitian.....	40
BAB VI PENUTUP	
6.1 Kesimpulan	43
6.2 Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN	47

DAFTAR GAMBAR

No.	Judul	Halaman
Gambar 2.1	Diagram Sebaran Data Korelasi	10
Gambar 3.1	Kerangka Teori Penelitian	14
Gambar 3.2	Diagram Alir Langkah Penelitian.....	15
Gambar 4.1	Diagram Alir Metodologi Penelitian	19
Gambar 4.2	<i>Flowchart</i> Pendekatan Korelasi	25
Gambar 4.3	<i>Flowchart</i> Klasifikasi <i>K-Nearest Neighbor</i>	27
Gambar 5.1	Grafik Hasil Klasifikasi Setelah dan Sesudah Perbaikan	38

DAFTAR TABEL

No.	Judul	Halaman
Tabel 4.1	Kisi – Kisi Kuisisioner.....	22
Tabel 4.2	Kriteria Tingkat Korelasi	26
Tabel 5.1	Data Hasil Kuisisioner Masuk Dalam Kriteria	29
Tabel 5.2	Tingkat Korelasi Variabel Cadangan Hobi dan Cita – Cita	30
Tabel 5.3	Tingkat Korelasi Variabel Cadangan Pekerjaan Orangtua dan Tes Kesehatan	30
Tabel 5.4	Klasifikasi Jenis Variabel Cadangan Hobi	31
Tabel 5.5	Klasifikasi Jenis Variabel Cadangan Cita - cita	32
Tabel 5.6	Klasifikasi Jenis Variabel Cadangan Pekerjaan Orangtua	32
Tabel 5.7	Daftar Kasus Tidak Mengalami <i>Missing Value</i>	33
Tabel 5.8	Data <i>Missing Value</i> Pada Variabel 3 dan Variabel 4 Kelas X	35
Tabel 5.9	Data <i>Missing Value</i> Pada Variabel 3 dan Variabel 4 Kelas XI	36
Tabel 5.10	Data <i>Missing Value</i> Sudah diganti Data Cadangan Kelas X	36
Tabel 5.11	Data <i>Missing Value</i> Sudah diganti Data Cadangan Kelas XI	37
Tabel 5.12	Perbandingan dan Tingkat Akurasi Nilai k	38
Tabel 5.13	Klasifikasi K-NN Dengan Nilai k.....	39

RINGKASAN

Novta Dany’el Irawan, Jurusan Teknik Elektro, Fakultas Teknik Universitas Brawijaya, November 2017, Analisis Dan Implementasi Pendekatan Korelasi Untuk Perbaikan *Missing Value* Dengan Metode *K-Nearest Neighbor(K-NN)*, Dosen Pembimbing : Wijono dan Onny Setyawati. *Missing value* sering terjadi dalam metode klasifikasi disebabkan karena informasi tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada, hal ini akan menyebabkan menurunnya keakuratan dan kualitas data pada saat data diolah. Mengganti variabel kosong dengan estimasi dari beberapa nilai tidak dapat langsung dilakukan jika variabel kosong. Diperlukan pendekatan spesifik untuk mengetahui korelasi antara variabel kosong (*missing value*) dengan variabel pengganti atau cadangan. Menentukan jurusan bagi siswa SMA, murid dan orang tua mereka sering mengabaikan untuk mengisi data yang lengkap. Data mengalami *missing value* dapat diatasi dengan menambahkan data variabel cadangan yang diperoleh dari hasil pengukuran menggunakan pendekatan korelasi dengan metode klasifikasi K-Nearest Neighbor (K-NN).

Pendekatan korelasi dilakukan karena peneliti perlu mengetahui ketersediaan dan kekuatan korelasi variabel yang berkaitan dengan objek. Pendekatan korelasi bertujuan untuk mengetahui bukti adanya korelasi antara variabel dan untuk mendapatkan korelasi signifikan atau tidak signifikan. Metode klasifikasi yang digunakan dalam penelitian ini adalah K-NN, karena termasuk metode klasifikasi dengan konsistensi kuat, menghitung hubungan antara kasus baru dan kasus lama berdasarkan nilai k dan tetangga terdekatnya.

Mengamati hasil kuesioner dilakukan untuk meneliti nilai korelasi data cadangan, hal ini termasuk dalam kategori data cadangan yang layak. Hasil percobaan pada k nilai metode K-NN dengan menggunakan k value = 3, 5, 7, 9, dan 11 dari 500 data dan 31 pola kasus lama, menunjukkan bahwa akurasi yang tinggi 97% pada nilai k = 5.

Kata Kunci: *missing value, pendekatan korelasi, k-nearest neighbor (K-NN)*

SUMMARY

Novta Dany'el Irawan, Department of Electrical Engineering, Engineering Faculty, Brawijaya University, November 2017, *Analysis and Implementation of Correlation Approach to solve Missing Value problem with K-Nearest Neighbor(K-NN)*, Advisor: Wijono and Onny Setyawati

Missing value often occurred in classification method is caused by information on unavailable, and unaffordable object, or the unexisting information, that can lead to the depletion of data accuracy and quality during data processing. Changing empty variable using estimation of values cannot be directly conducted if the variable is empty. It needs specific approach to find correlation of empty variable (missing value) with substitute variable or spare. To determine major subject for Senior High Schoolers, pupils and their parents often ignore to fill the completely data. Data with missing value can be solved by adding spare variable data obtained from the result of measurement using correlation approach with K-Nearest Neighbor (K-NN) classification method.

Correlation approach is conducted since the researcher needs to find out the availability and the strength of variable correlation related to the object. The correlation approach aimed to find the proof on the existence of correlation between variables and to obtain the significant or insignificant correlation. The classification method used in this study is K-NN because it includes classification method with strong consistency, by calculating the relation between new case and old case based on k value and its nearest neighbor.

Observing the result of questionnaires are conducted to scrutinize correlation value of spare data, it is included in feasible category of spare data. Experimental results on k value of K-NN method by using k value = 3, 5, 7, 9, and 11 from 500 data and 31 old cases pattern, show the high accuracy of 97% at k value = 5.

Keywords: *missing value, correlation approach, k-nearest neighbor (K-NN)*